DOCUMENT RESUME

ED 210 291 TM 810 934

AUTHOR Gold, Robert S.
TITLE Measurement of the Perceived Characteristics of Evaluation Recommendations.
PUB DATE [79]
NOTE 19p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Change Agents; Evaluation; *Evaluators; *Measurement Techniques; *Reliability; Semantic Differential
IDENTIFIERS *Evaluation Reports; *Evaluation Utilization; Maine Health Education Resource Center

ABSTRACT

A study was conducted on the measurement of the characteristics of innovation. The results or specific recommendations of evaluation studies may be assessed for their likelihood of adoption and implementation based on these characteristics. As the recommendations are perceived more positively, the level of utilization should increase. The semantic differential technique developed by Osgood, Succi, and Tannenbaum was used to measure the characteristics in this study. The Statistical Package for the Social Sciences Reliability sub-program was used to assess the internal consistency of the scales to identify which recommendations are positively perceived. With this knowledge, an evaluator/change agent may be better able to judge where efforts should be directed. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED210291

Measurement of the Perceived Characteristics of Evaluation Recommendations

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R.S. Gold
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 534

Robert S. Gold, Ph.D., Dr. P.H., Associate Professor, Department of Health Education, Southern Illinois University, Carbondale, Illinois 62901

A great many definitions exist for evaluation. Few totally agree on what evaluation is and what its primary purposes are. There is, however, a thread that runs through most definitions of evaluation - that is, the major purpose of evaluation is to serve as input in the decision-making process (Merriman, 1972; Provus, 1972; Weiss, 1972a; and Patton, 1978).

The issue over utilization of evaluation results has been receiving increasing coverage recently. In spite of the fact that knowledge utilization studies have proliferated to more than 20,000 titles in the last 20 years, many will argue that few studies look at evaluation research as a source of such knowledge.

Davis and Salasin (1975) found that only 2.5 percent of the references in the utilization literature were related to evaluation, while only 5.0 percent of published evaluation studies mention utilization of results. Most of the studies that do examine the utilization of results from evaluation studies report low levels of utilization (Davis & Salasin, 1977). Patton (1978) believes that our definition of utilization has been too narrow in the past and argues that when the difinition is expanded to include any legitimate use of results, utilization patterns increase dramatically.

While a great deal of interest in this areas has been expressed, a continuing problem has been the lack of knowledge of the specific determinants of the utilization of evaluation results, the extent of their use, and the

absence in the literature of empirical investigations of those determinants (Weiss, 1972b). Though there are some studies on the determinants of utilization of evaluation results (Bernstein & Freeman, 1975; Brennan, 1976; Cohen, 1977; Patton et al., 1979), most of the empirical literature is generalized to the area of dissemination of knowledge and/or innovations and subsequent adoption of innovation (e.g. Fleigel & Kivlin, 1966; Becker, 1970a, 1972b; Blum, 1964), or evaluation of change efforts in organizational de-elopment (e.g. Hage & Aiken, 1967; Hage & Aiken, 1970; Jenks, 1970; Kelly, 1976; Killmann & Herden, 1976;, Kirchhoff, 1977; Lawler, 1977).

A great deal of model building has been done in the social sciences in an attempt to explain the relationship between large arrays of variables and utilization of knowledge from some source. Greer (1977) provides an excellent reivew of the literature available in this area. A relatively recent approach to the study of utilization of evaluation results has been the development of aggregate measures of organizational readiness for change, with Davis (1971) providing one of the prototypes with A VICTORY. Most of these aggregate measures provide much the same type of information as is provided by the structural theorists, however, A VICTORY provides an assessment of staff perceptions of organizational readiness to adopt innovations.

A more recent model developed by Seigel and Kaemmerer (1978) contained

three factors associated with organizational readiness to change or innovate: 1) a factor labeled support for creativity; 2) a second factor called tolerance of differences; and 3) a factor called personal commitment. However, the work by Rogers and Shoemaker (1971) on the characteristics of the innovations themselves, provide a structure for the classification of innovations that Becker (1970a) felt vital to unmasking spurious relationships between predictors and rate of adoption.

The current paper describes an initial study on the measurement of the characteristics of innovation based on the work by Rogers and Shoemaker (1971). In this approach, the results (i.e. specific recommendations) of evaluation studies may be assessed for their liklihood of adoption and implementation based on these characterisitcs. Knowledge of these perceived characteristics may then be helpful when the evaluator takes on the change agent role recommended by Davis and Salasin (1975). According the Davis (1971), Rogers and Shoemaker (1971), Glaser (1976) and Zaltman and Duncan (1977), it would be expected that as the recommendations themselves are perceived more positively, the level of utilization would increase.

The Setting

The current study was conducted during a year long evaluation project at the Maine Health Education Resource Center (HERC), a project in the Divi-

sion of Health at the University of Maine at Farmington. During its first three years of operation, HERC's programmatic efforts were directed primarily at those health areas in which individual life style changes and wiser use of the health care system were thought to be beneficial. A final report from a comprehensive evaluation of HERC was completed during the summer of 1979. That final report contained the results of the evaluation of one of the units of HERC, the Unit for Demonstration Projects, and in an Executive Summary seven primary recommendations were made (Attachment I). The current study focused on the measurement of the perceived characteristics of these recommendations.

Methods

The characteristics measured in the current study were much like those recommended by Rogers and Shoemaker (1971), including the relative advantage of the recommendations, their compatibility with both the values and needs of HERC, their complexity and trialability.

Instruments

The semantic differential (SD) technique, developed by Osgood, Succi and Tannenbaum (1957) was selected as an approach to the measurement of these characteristics because 1) it is an efficient means of getting quantifiable data not only for direction but also intensity of attitudes towards a concept;

2) it provides a comprehensive image; 3) it represents a standardized technique; 4) it is easily repeatable and quite reliable, therefore useful for monitoring; 5) it avoids stereotyped responses; and 6) it eliminates some of the problems of question phrasing (Mindak, 1956: 368).

The scale items were used individually and as a composite score to get a measure of the connotations of each of the recommendation on each of the dimensions identified in the literature. Each of the recommendations was listed as a concept followed by a list of the scale items. The same scale items were used for each concept and in the same order. The polarity of some of the scales were reversed so that digit preference might be reduced.

A mean for each scale was calculated by summing across all participants for each recommendation. Composite means were calculated as the simple sum of the scale items for each recommendation. Figure 1 contains a sample of the format used.

Insert Figure 1 About Here

Assessment of the Reliability of the Scales

The Statistical Package for the Social Sciences (SPSS) Reliability subprogram was used to assess the internal consistency of the scales based on the

calculation of Cronbach'a alpha. An alpha approaching 1.0 indicates a scales with a high degree of internal consistency, while an alpha approaching 0 indicates greater divergence among the items of the scale. An alpha coefficient of 0 suggests that the items making up the scale are not measuring the same attribute. To facilitate the search for divergent or inconsistent scale items, item-total correlation coefficients may be computed and the alpha coefficient may be recalculated with each scale item removed from the analysis in turn. The resulting coefficients may then be examined directly for the influence of each item. If the correlation of the item to the total is low, or if the alpha coefficient increases when a scale item is removed, that item should be reviewed substantively to determine if it is addressing the same question or attribute as the remaining items. Each of the seven semantic differential scales was analyzed separately for their estimated reliability.

Sample

The invited sample consisted of all personnel working at the HERC (in Farmington) (n = 15), the HERC Advisory Board (n = 25) and the 1978-1979 External Evaluation Panel (n = 6) for a total of 46 persons. Each of these individuals was invited to participate after the nature of the study was explained in writing and they had an opportunity to have any of their questions

answered. Participation was voluntary and participants were asked to sign a consent form.

Results

Sample

The complete study sample of 24 was composed of 13 staff members, 7 advisory board members and 4 evaluation panel members. The staff respondents represented 86.67 percent of the toal staff invited to participate (n = 15). Initially, the participation of 25 advisory board members was sought. However, because of board turnover, the seven participants from the board came from an available pool of approximately 15 persons. This resulted in a 46.67 percent response rate, although the most rigid notion of response rate would result in a figure of 28.00 percent (7 accepting from 25 invited). Similarly, one resignation occurred among the 1978-1979 evaluation panel members, so the pool of possible participants was reduced to five from that source. The response rate in this case was 80.0 percent. Overall, the participating sample represented 68.57 percent of the available pool of subjects.

Scale Reliability

Each of the seven recommendations was made into a semantic differential scale containing 10 bi-polar items. The order and polarity of the pairs remained identical from scale to scale, and in each case, the positive pole

was assigned a value of 1 and the negative pole a value of 7. The reliability of each of the seven scales was assessed separately. Because the SPSS reliability package algorithm uses a listwise deletion of cases with missing data, the results reported here are based on 21 complete response sets.

Table 1 contains a summary of the overall reliability assessments for the seven semantic differential scales. The recommendation perceived as most positive based on the simple sum of the 10 item ratings was recommendation number 4 (dissemination of findings) with a composite mean of 15.71. In decreasing order, those recommendations perceived less positively overall were numbers 5 (long-range program planning, mean = 15.90), 2 (effective management procedures, mean = 16.24), 7 (commitment to evaluation, mean = 17.19), 1 (re-examine original HERC goals, mean = 18.24), 3 (undergraduate program development, mean = 18.57), and 7 (data-based evaluation, mean = 21.43).

The alpha coefficients from each of the full 10-item scales ranged from a low of .6602 (scale 2: effective management procedures) to a high of .8991 (scale 1: re-examine original HERC goals). Only two scales (2 and 3) had alpha coefficients less than .80.

A visual examination of both the inter-item correlation matrices and the summary item-total statistics indicated that the item representing the reversible-irreversible domain consistently had the lowest inter-item cor-

relation coefficient, as well as the only negative coefficients. For each recommendation this item had the lowest item-total correlation and resulted in the maximum increase in the magnitude of the alpha for that scale when it was removed. This would indicate that this item was most discordant with the remaining items on that individual scale, as was clearly reflected in the low or negative inter-item correlations between this item and virtually all other items for each scale. However, as each set was constituted with the full 10 items, the reliability (represented by internal consistency) appeared to be sufficient for use as a scale.

Insert Table 1 About Here

Discussion

While Heise (1969) raises some issues about the use of SD scales because of some problems with its metric characteristics, these should not preclude the use of SD as proposed in the current paper. Here the concern is not for an absolute measure, but rather a relative measure of the perceptions of certain characteristics of evaluation recommendations. The results of the reliability analysis indicate sufficient internal consistency of the scales to identify which recommendations are more or less positively perceived. With

this knowledge, an evaluator/change agent may be better able to judge where efforts should be directed. Scaling such as this could perhaps best be used as an adjunct to measures of organizational readiness to change to provide a clearer picture of the liklihood of adoption and implementation of evaluation recommendations.

Selected References

Becker, M. H. Factors affecting diffusion of innovations among health professionals. American Journal of Public Health, 1970a, 60(2), 294-304.

Becker, M. H. Sociometric location and innovativeness: Reformulation and extension of the diffusion model. American Sociological Review, 1970b, 35, 267-282.

Bernstein, I. N., and Freeman, H. E. Academic and entrepreneurial research: The consequences of diversity in federal evaluation studies. New York: Russel Sage Foundation, 1975.

Blum, R. H., and Downing, J. J. Staff response to innovation in mental health service. American Journal of Public Health, 1964, 54, 1230-1240.

Brennan, N. J. Variation in the utilization of evaluation research in decision making. Unpublished PhD dissertation, University of Minnesota, 1976.

Cohen, L. H. Factors affecting the utilization of mental health evaluation research findings. Professional Psychology, 1977, 8(4), 526-534.

Davis, H. R. A checklist for change In National Institute of Mental Health, A manual for research utilization. Washington, D. C.: Government Printing Office, 1971.

Davis, H. R., and Salasin, S. E. Applied social research: In combat with waste and suffering. In M. Guttentag, and S. Saar (eds.) Evaluation studies review annual: Volume 2. Beverly Hills: Sage Publications, 1977.

Davis, H. R., and Salasin, S. E. The utilization of evaluation. In E. L. Struening, and M. Guttentag (eds.) Handbook of evaluation research: Volume 1. Beverly Hills: Sage Publications, 1975.

Fliegel, F. C., and Kivlin, J. E. Attributes of innovations as factors in diffusion. American Journal of Sociology, 1966, 72(3), 235-248.

Glaser, E. M. Putting knowledge to use: A distillation of the literature regarding knowledge transfer and change. Rockville, Maryland: National Institute of Mental Health, 1976.

Greer, A. L. Advances in the study of diffusion of innovation in health care organizations. Milbank Memorial Quarterly, 1977, 55(4), 505-532.

Hage, J., and Aiken, M. Program change and organizational properties: A comparative analysis. American Journal of Sociology, 1967, 72(5), 503-519.

Hage, J., and Aiken, M. Social change in complex organizations. New York: Random House, 1970.

Heise, D. R. Some methodologic issues in semantic differencial research. Psychological Bulletin, 1969, 72(6), 406-422.

Jenks, R. S. An action research approach to organizational change. Journal of Applied Behavioral Science, 1970, 2, 131-150.

Kelly, G. Seducing the elites: The politics of decision making and innovation in organizational networks. Academy of Management Review, 1976, 1(2), 66-74.

Kilmann, R. H., and Herden, R. P. Towards a systematic methodology for evaluating the impact of interventions on organizational effectiveness. Academy of Management Review, 1976, 1(2), 87-98.

Kirchhoff, B. A. Organization effectivenss measurement and policy research. Academy of Management Review, 1977, 2(3), 347-355.

Lawler, E. E. III. Adaptive experiments: An approach to organizational behavior research. Academy of Management Review, 1977, 2(4), 576-585.

Merriman, H. O. Evaluation of planned change at the local educational level. In P. A. Taylor and D. M. Cowley (eds.) Readings in curriculum evaluation. Dubuque, Iowa: W. C. Brown, 1972.

Mindak, W. A. A new technique for measuring effectiveness. Journal of Marketing, 1956, 20, 367-378.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. The measurement of meaning. Urbana, Illinois: University of Illinois Free Press, 1957.

Patton, M. Q. Utilization focused evaluation. Beverly Hills: Sage Publications, 1979.

Patton, M. Q., Grimes, P. S. Guthrie, K. M., Brennan, N. J., French, B. D., and Blyth, D. A. In search of impact: An analysis of the utilization of federal health evaluation research. In T. D. Cook, M. L. Del-Rosario, K. M. Hennigan, M. M. Mark, and W. M. K. Trochim (eds.) Evaluation studies review annual: Volume 3. Beverly Hills: Sage Publications, 1978.

Provus, M. M. The discrepancy evaluation model. In P. A. Taylor and D. M. Cowley (eds.) Readings in curriculum evaluation. Dubuque, Iowa: W. C. Brown, 1972.

Rogers, E. M., and Shoemaker, F. F. Communication of innovations: A cross-cultural approach. New York: The Free Press, 1971.

Seigel, S. M., and Kaemmerer, W. F. Measuring the perceived support for innovation in organizations. Journal of Applied Psychology, 1978, 63(5),553-562.

Weiss, C. H. Between the cup and the lip. Evaluation, 1973a, 1(2), 49-55.

Weiss, C. H. Evaluation research: Methods of assessing program effectiveness. Englewood Cliffs: Prentice Hall, 1972b.

Zaltman, G., and Duncan, R. Strategies for planned change. New York: John Wiley & Sons, 1977.

Table 1

Summary Statistics For the Semantic Differential Scales

Scale	Mean	Alpha	Standardized Item Alpha
Re-examine original goals	18.24	.8991	.9103
Effective management procedures	16.24	.6602	.8028
Undergraduate program development	18.57	.7756	.8562
Dissemination of findings	15.71	.8101	.8666
Long range program planning	15.90	.8161	.8823
Data-based evaluation	21.43	.8975	.9198
Commitment to evaluation	17.19	.8178	.8946

Figure 1

Sample Semantic Differential Scale Used
To Assess The Characteristics of Each Recommendation

STATEMENT OF RECOMMENDATION

DESIRABLE	___:___:___:___:___:___:___	UNDESIRABLE
FEASIBLE	___:___:___:___:___:___:___	UNFEASIBLE
IMPORTANT	___:___:___:___:___:___:___	UNIMPORTANT
IRRELEVANT	___:___:___:___:___:___:___	RELEVANT
ADVANTAGEOUS	___:___:___:___:___:___:___	DISADVANTAGEOUS
IRREVERSIBLE	___:___:___:___:___:___:___	REVERSIBLE
TIMELY	___:___:___:___:___:___:___	UNTIMELY
NOT COMPATIBLE WITH HERC VALUES	___:___:___:___:___:___:___	COMPATIBLE WITH HERC VALUES
NOT COMPATIBLE WITH HERC NEEDS	___:___:___:___:___:___:___	COMPATIBLE WITH HERC NEEDS
CAN BE PARTIALLY IMPLEMENTED	___:___:___:___:___:___:___	CANNOT BE PARTIALLY IMPLEMENTED

Attachment I: Evaluation Recommendations

1. The HERC goals stated in the original grant proposal should be re-examined critically. That examination should result in a restatement of goals that more accurately reflects the current mission and purpose of HERC.

2. Lines of authority, responsibility, and communication pertaining to HERC should be improved, and specific procedures for effective management should be developed.

3. The University of Maine at Farmington should continue moving toward establishment of a high quality undergraduate program in community health education as an important partner of HERC.

4. HERC should develop appropriate methods of sharing its findings within the state and beyond.

5. HERC should develop detailed long-range and short-range plans for comprehensive program development.

6. The data-based evaluation initiated in 1978-1979 should be continued and expanded in 1979-1980.

7. A commitment to the continued evaluation of the demonstration projects should be expressed in the contractual agreements between HERC and participating project sites.